New Phytologist

Bridging the genotype-phenotype gap for a Mediterranean pine by semi-automatic crown identification and multispectral imagery

Journal:	*New Phytologist*
Manuscript ID	NPH-MS-2020-32375.R1
Manuscript Type:	MS - Regular Manuscript
Date Submitted by the Author:	11-Jun-2020
Complete List of Authors:	Santini, Filippo; University of Lleida School of Agricultural and Forestry Engineering, Departament de Producció Vegetal i Ciència Forestal Kefauver, Shawn; Universitat de Barcelona Facultat de Biologia, Biologia Evolutiva, Ecologia i Ciències Ambientals Araus, José; Universitat de Barcelona, Facultat de Biologia, Biologia Evolutiva, Ecologia i Ciències Ambientals Resco de Dios, Victor; University of Lleida School of Agricultural and Forestry Engineering, Departament de Producció Vegetal i Ciència Forestal Martín García, Saray; föra forest technologies sll, Campus Duques de Soria Grivet, Delphine; Center of Forest Research (CIFOR-INIA), Unit of Forest Genetics Voltas, Jordi; University of Lleida School of Agricultural and Forestry Engineering, Departament de Producció Vegetal i Ciència Forestal
Key Words:	Aleppo pine, Genomics, GWAS, Remote Sensing, SNPs, UAV

Bridging the genotype-phenotype gap for a Mediterranean pine by semi-automatic crown identification and multispectral imagery

Filippo Santini[1,2*], Shawn Carlisle Kefauver[3,4], José Luis Araus[3,4], Víctor Resco de Dios[1,2,5], Saray Martín García[6], Delphine Grivet[7], Jordi Voltas[1,2]

[1]Joint Research Unit CTFC - AGROTECNIO, Av. Alcalde Rovira Roure 191, E-25198 Lleida, Spain;
[2]Department of Crop and Forest Sciences, University of Lleida, Av. Alcalde Rovira Roure 191, E-25198 Lleida, Spain;
[3]AGROTECNIO (Center for Research in Agrotechnology), Av. Alcalde Rovira Roure 191, E-25198 Lleida, Spain;
[4]Integrative Crop Ecophysiology Group, Plant Physiology Section, Faculty of Biology, University of Barcelona, E-08028 Barcelona, Spain;
[5]School of Life Science and Engineering, Southwest UNiversity of Science and Technology, Mianyang, China;
[6]föra forest technologies, c/Eduardo Saavedra 38, E-42004 Soria, Spain;
[7]Department of Forest Ecology and Genetics, Forest Research Centre, INIA, Carretera A Coruña km 7.5, E-28040 Madrid, Spain

*Corresponding author: Filippo Santini
Department of Crop and Forest Sciences
ETSEA – University of Lleida
Av. Alcalde Rovira Roure 191
E-25198 Lleida, Spain
Tel. +34 973 702622
E-mail: filippo.santini@pvcf.udl.cat

Total word count: 6526

Introduction word count: 1042

Materials and Methods word count: 2375

Results word count: 929

Discussion word count: 2180

Number of figures: 4

Colour figures: Fig.1, Fig.2 and Fig.4

Number of Tables: 4

Supplementary information: 3 tables; 7 figures; 4 Methods

SUMMARY

- Progress in high-throughput phenotyping and genomics provides the potential to understand the genetic basis of plant functional differentiation. We developed a semi-automatic methodology based on Unmanned Aerial Vehicle (UAV) imagery for deriving tree-level phenotypes followed by genome-wide association study (GWAS).
- A RGB-based point cloud was used for tree crown identification in a common garden of *Pinus halepensis* in Spain. The crowns were combined with multispectral and thermal orthomosaics to retrieve growth traits, vegetation indices and canopy temperature. Thereafter, GWAS was performed to analyse the association between phenotypes and genomic variation at 235 Single Nucleotide Polymorphisms (SNPs).
- Growth traits were associated with 12 SNPs involved in cellulose and carbohydrate metabolism. Indices related to transpiration and leaf water content were associated with six SNPs involved in stomata dynamics, and indices related to leaf pigments and leaf area were associated with 11 SNPs involved in signalling and peroxisomes metabolism. About 16% to 20% of trait variance was explained by combinations of several SNPs, indicating polygenic control of morpho-physiological traits.
- Despite a limited availability of markers and individuals, this study is a successful proof-of-concept for the combination of high-throughput UAV-based phenotyping with cost-effective genotyping to disentangle the genetic architecture of phenotypic variation in a widespread conifer.

Keywords: Genomics; GWAS; *Pinus halepensis*; Remote Sensing; SNPs; UAV

INTRODUCTION

The rapid development of genotyping and phenotyping technologies is narrowing down the knowledge gap between genomics and phenomics (Houle *et al.*, 2010; Großkinsky *et al.*, 2015). Hundreds of individuals can be characterised with an ever-growing number of genetic markers, covering large parts of the genome. As a consequence, the amount of genome-wide association studies (GWAS) seeking to understand the genetic basis underlying phenotypic differentiation in plants has increased exponentially (Lobos *et al.*, 2017). Common garden experiments, where individuals from contrasting geographic origins grow under the same environmental conditions, are often used in ecological studies to understand to what extent individual differences are controlled by genetic effects (McKown *et al.*, 2014; Baison *et al.*, 2019). Combining accurate and rapid phenotyping with genomic scans in common gardens can therefore result in an effective method to study the genetic basis of adaptation in plant species (de Villemereuil *et al.*, 2016).

Efforts towards understanding the genetic basis of adaptation in forest tree species have been traditionally limited by a lack of laboursaving phenotyping techniques, as well as by a restricted amount of molecular information available for non-model organisms. Genome-wide sets of molecular markers are needed in order to gain information on the number of loci contributing to the genetic architecture of complex traits (Grattapaglia & Resende, 2011). Because most traits are controlled by many loci with small effects, a high density of loci is required to achieve a solid accounting of phenotypic variation (White *et al.,* 2007; Savolainen *et al.*, 2007). Once available for only a few model organisms, sets of molecular markers putatively under selection are now accessible for non-model species such as forest trees (Khan & Korban, 2012; Jaramillo-Correa *et al.*, 2015).

Furthermore, the development of suitable phenotyping approaches is strongly limited by the costs associated with manual phenotypic measurements of anatomical and physiological traits. These costs derive from the long life-cycles of forest trees and the considerable dimensions of adult individuals (Ludovisi *et al.*, 2017). As an alternative, remote sensing using Unmanned Aerial Vehicles (UAV) imagery is emerging as an effective high-throughput phenotyping tool to indirectly infer variation in common garden experiments of forest trees (Santini *et al.*, 2019a). Based on specific wavelengths, UAV-based multispectral imagery allows the calculation of vegetation indices, which can be used as proxies of physiological performance. This approach has been used to obtain information about traits such as crown characteristics, leaf anatomy and content of photosynthetic pigments in leaves (Santini *et al.*, 2019a; 2019b). Additionally, water status and total transpiration can be inferred through thermal images in trees growing in field

experiments (Ludovisi *et al.*, 2017; Santini *et al.*, 2019a). To complement the aforementioned information, innovative techniques of photogrammetry can be applied to RGB (red, green, blue) images to obtain 3D reconstructions of trees. These techniques aim at the automatic identification of single trees in forests and, also, at estimating growth parameters such as tree diameter or height (Nevalainen *et al.*, 2017). Therefore, high-resolution remote sensing data coupled with efficient algorithms for automatic crown identification represent a potentially powerful approach for fast and accurate individual tree phenotyping (Wallace *et al.*, 2016; Santini *et al.*, 2019a). In previous studies, we developed a UAV-derived phenotyping methodology that successfully detected potentially adaptive variation in leaf area, crown architecture, transpiration rate and photosynthetic pigments in *Pinus halepensis* Mill. (Santini *et al.*, 2019a) and *Pinus nigra* Arnold (Santini *et al.*, 2019b). Here, we further developed and combined this methodology with genomic information in a common garden of *P. halepensis* to evaluate the potential of linking novel phenotyping technology with high-throughput genotyping to identify candidate genes involved in adaptive differentiation for a widespread forest species.

Aleppo pine (*Pinus halepensis*) is a drought-resistant species and the most common conifer across the Mediterranean basin. Common garden experiments have revealed a wide differentiation among populations in traits such as aerial growth (Schiller & Atzmon, 2009; Voltas *et al.*, 2018), phenology (Klein *et al.*, 2013), water uptake patterns (Voltas *et al.*, 2015), hydraulic conductivity (Tognetti *et al.*, 1997) and reproductive effort (Climent *et al.*, 2008). With the goal of unveiling the genetic basis underlying this differentiation, which remains largely unexplored, a set of Single Nucleotide Polymorphism (SNP) markers was developed by Pinosio *et al.* (2014). Ruiz-Daniels *et al.* (2018; 2019) used these markers to genotype natural populations both range-wide and at local scale, and detected several SNPs putatively associated with adaptation to local conditions. A limitation of these approaches derives from the complex neutral genetic structure of this species (i.e. unrelated to local adaptation), which has been shaped by past population dynamics (Serra-Varela *et al.*, 2017; Ruiz Daniels *et al.*, 2018). The strong genetic structure separating Western and Central-Eastern Mediterranean populations, with further subdivisions within each group (Ruiz Daniels *et al.*, 2018), is prone to generate false positive associations related to non-adaptive differentiation among populations with divergent phenotypes, making the identification of loci associated with phenotypic changes challenging (Yu *et al.*, 2006). Previous analyses performed in natural populations can be complemented with common-garden information for the purpose of identifying SNPs putatively involved in relevant adaptive traits for this pine species by controlling for the effects of environmental variation and neutral structure.

In this study, we used *P. halepensis* populations grown in a common garden as a model system to provide a novel approach for high-throughput phenotyping of a suite of traits related to tree morphology and physiology, which are (in most cases) seldom considered in genotype-phenotype association studies. We aimed at deepening into the genetic control of these traits by combining phenotyping through semi-automatic crown identification and multispectral imagery with genotyping at individual level. We hypothesized that, for a species encompassing a large functional variability in characteristics related to water and carbon economy, the genetic basis of physiological and morphological variability is identifiable, at least in part, through high-throughput UAV imagery. Specifically, the objectives of this study were to: 1) develop a high-throughput methodology for deriving tree-level information of growth traits, together with vegetation indices and canopy temperature; 2) identify putative genes underlying the variation in such traits, test for multi-genic control of phenotypic variation and estimate the proportion of phenotypic variation explained by genotypes; and 3) provide an insight into the potential of combining UAV imagery with genotypic information in disentangling adaptive variation in forest species.

MATERIALS AND METHODS

Study site and plant material

The study was carried out on adult individuals of Aleppo pine growing in a common garden experiment in Altura (39°49′29″N, 00°34′22″W, 640 m a.s.l.; Castellón province, eastern Spain) (Supporting Information Fig. S1). Seeds from 56 natural populations of *P. halepensis* were collected in 1995 and planted in a forest nursery in Spain (Table S1, Fig. S2). These populations cover a large part of the species' distribution range and they are representative of the wide variety of environmental conditions in which *P. halepensis* is currently found. For each population, open-pollinated seeds were collected from 20 to 30 trees, spaced at least 100 m apart, and subsequently bulked into population seedlots. This sampling strategy was implemented to minimize genetic relatedness among the half-sib families and, also, to collect a representative amount of intra-population variability. In 1997, one-year old seedlings were planted systematically (2.5 × 2.5 m spacing) at the study site in experimental units consisting of linear plots of four individuals. Four replicates were established following a Latinised row-column design for a total of 896 individuals (16 per population) tested in the trial. Height and diameter at breast height (DBH) were ground-measured for all trees in 2013 (at age 16).

UAV flights

Flights were performed at noon on 26th July 2016 (at age 19 years) using an octocopter (Mikrokopter OktoXL; HiSystems GmbH, Moormerland, Germany) flying under remote control at approximately 100 m a.g.l. (above ground level). The trees' crown cover fraction at the trial was 65% (Santini *et al.*, 2019a). A significant crown overlapping was observed between neighbouring trees (Fig. S1). To ensure high quality image registration and an overlapping between pictures of approximately 80%, forward motion was kept to less than 5 m s^{-1} during flights. The duration of the flights was 10 minutes and the total area covered was *ca.* (circa) 0.6 ha, corresponding to the extension of the common garden.

The UAV was equipped with three different cameras that were mounted on a two-servo gimbal (MK HiSight SLR2; HiSystems GmbH, Moormerland, Germany) nadir-looking in consecutive flights. First, a multispectral camera (MCA12; Tetracam Inc., Chatsworth, CA, US) was operated to capture 15.6-megapixel calibrated-reflectance images in 10 wavelengths (450 ± 40, 550 ± 10, 570 ± 10, 670 ± 10, 700 ± 10, 720 ± 10, 840 ± 10, 860 ± 10, 900 ± 20, 950 ± 40 nm) in the visible and near infrared (NIR) regions of the spectrum. Second, RGB images were obtained using a Mirrorless Interchangeable Lens Camera (Lumix GX7; Panasonic, Osaka, Japan). Finally, a FLIR thermal camera (Tau2 640; FLIR Systems, Nashua, NH, USA) was employed for the acquisition of thermal images. Further details on camera calibration, validation of reflectance data and thermal data retrieval are reported in Methods S1 and S2.

Image processing and crown identification

In previous studies, we used a manual approach to derive plot-level phenotypic data from UAV imagery in this same trial (Santini *et al.*, 2019a). Here, we developed a workflow, schematised in Fig. 1, for the semi-automatic acquisition of tree-level phenotypic data. Two orthomosaics (one multispectral and one thermal) were obtained from raw multispectral and thermal images using the Agisoft PhotoScan Professional software (Agisoft LLC, St. Petersburg, Russia). The spatial resolution of the multispectral and thermal orthomosaics were 7 cm^2 and 29 cm^2 per pixel respectively. The geographic coordinates of four ground control points were obtained with 1-meter accuracy using a GPS device (Juno 5B, Trimble Inc., Sunnyvale, USA) and were used for georeferencing the orthomosaics.

RGB images were analysed through structure-from-motion (SfM) photogrammetry in the software Agisoft PhotoScan Professional to obtain a georeferenced 3D dense point cloud of the common garden, in which the height of each point was expressed in meters above sea level. The software FUSION (McGaughey, 2014) was employed to classify vegetation and soil points of

the dense point cloud, and a digital terrain model (DTM) was obtained thereafter (Fig. 2A). Subsequently, the dense point cloud and the DTM were combined in FUSION to obtain a normalized point cloud, in which the height of each point is expressed in meters above ground. Finally, we obtained the canopy height model (CHM) from the normalized point cloud using the function grid_canopy implemented in the package lidR (Roussel *et al.,* 2018) of the R environment (R Core Team, 2019) (Fig. 2B).

The CHM was used to identify treetops using an algorithm based on local maximum filter (Popescu & Wynne, 2004) implemented in the function tree_detection in the R package lidR (Roussel *et al.,* 2018). The height of each treetop was retrieved from the CHM as an imagery-derived estimation of tree height (H_{UAV}). Finally, single crowns were segmented (i.e. outlined) in the CHM using the function mcws in the ForestTools package (Plowright, 2018), which performs a watershed segmentation (Meyer & Beucher, 1990) guided by the locations of the treetops (Fig. 2B). Tree crown area was calculated as the projection of the outlined crown on the ground. Crown area was considered as a surrogate of DBH (Lockhart *et al.*, 2005) (Fig. S3).

The accuracy of tree identification was visually checked in the software QGIS (version 3.6; QGIS Development Team, 2019) and the crowns of the misidentified trees were manually corrected (Fig. 2C, D). The obtained georeferenced crown shapes were used to identify the individual crowns from multispectral and thermal orthomosaics, that is, by using the crown shapes as template images corresponding to single trees, crowns were cropped from the orthomosaics (Fig. S3). Therefore, two images of the crown (one multispectral and one thermal) were retrieved for each of the 806 living trees of the common garden (see "Results" section) and used to calculate tree-level vegetation indices and canopy temperature.

Vegetation indices and canopy temperature

We calculated ten vegetation indices (Table 1) from multispectral images for each pixel of every single image, which corresponded to a particular tree crown. An average value was obtained afterwards for each of the indices per image (tree), resulting in individual tree data. Indices related to leaf area were calculated including the Normalised Difference Vegetation Index (NDVI), the Optimised Soil Adjusted Vegetation Index (OSAVI), the Re-normalized Difference Vegetation Index (RDVI) and the Enhanced Vegetation Index (EVI). Additionally, we also calculated the Modified Chlorophyll Absorption Reflectance Index (MCARI) and the Transformed Chlorophyll Absorption Ratio Index (TCARI). The latter indices are related to both leaf area and chlorophyll content (Daughtry, 2000). The ratio between TCARI and OSAVI (TCARI/OSAVI index) was calculated as a better estimate of chlorophyll content, i.e. free of

leaf area effects (Zarco-Tejada *et al.*, 2004). Furthermore, we calculated the Carotenoid Reflectance Index 2 (CRI2), the Anthocyanin Reflectance Index 2 (ARI2) and the Water Band Index (WBI) to investigate other photosynthetic pigments and water content in needles. Finally, we used thermal images to retrieve crown-level estimates of canopy temperature. A detailed description of vegetation indices is available in Methods S3).

Genotyping

A subset of 375 individuals belonging to 28 out of the 56 tested populations (Table S1, Fig. S2) were genotyped in a previous study using two sets of molecular markers (Pinosio *et al.*, 2014; Ruiz Daniels *et al.*, 2018). The first set consisted of eight nuclear simple sequence repeats (SSRs) loci, and was used to assess both population structure and relatedness between individuals (i.e. kinship matrix). A detailed description of sample collection and SSR amplification is reported in Ruiz Daniels *et al.* (2018).

The second set comprised 294 SNPs included in loci derived from transcriptomes of *P. halepensis* evaluated in their responses to fire, as well as from re-sequenced loci firstly identified in loblolly pine (*Pinus taeda*) and identified from previous studies as candidate involved in adaptation of the latter species (Pinosio *et al.*, 2014). This panel of markers thus comprised candidate genes related to wood anatomy, growth, phenology and, also, to a wide range of responses to abiotic stresses such as drought, cold and oxidative stress (Methods S4). This second set was used to identify genotype-phenotype associations.

Statistical analyses

Phenotypic data

To validate the image-derived estimations of growth traits, H_{UAV} and crown area were compared by means of simple correlations to ground-based measurements of height and DBH obtained in 2013. In order to remove the effects of heterogeneous growing conditions on phenotypic records, tree-level estimates of H_{UAV}, crown area, vegetation indices and canopy temperature were subjected to mixed-effects linear models. Analyses of variance (ANOVA) consisted of column and replicate as fixed terms and column by replicate interaction, row nested to replicate, and row nested to replicate by column interaction as random terms. The tree-level residuals of this model were individuals' phenotypic data which retained only genotypic variation (i.e., they were largely free of environmental effects, which were accounted for by trial design). This dataset was subsequently used for genotype-phenotype association analyses (see below), and was also used as input for a principal component analysis (PCA) with the objective of summarizing the

information retrieved by vegetation indices and to interpret tree-level relationships among traits. To this end, the loadings of H_{UAV}, crown area, vegetation indices and canopy temperature were plotted for the first two components of the PCA. Simple correlations were also calculated among phenotypic variables.

SSRs data

Individual genotypes at SSRs loci were used to calculate relatedness between individuals. For this purpose, a matrix of kinship between each pair of individuals was obtained from SSRs with the software SpaGeDi 1.3 (Hardy & Vekemans, 2015) using the kinship coefficient developed by Loiselle *et al.* (1995). In addition, the genetic structure of populations was inferred using the Bayesian clustering method implemented in STRUCTURE (Pritchard *et al.,* 2000). We ran STRUCTURE varying the numbers of possible genetic clusters (K) from one to ten, and each K was replicated ten times. Each run consisted of 1×10^5 burn-in iterations and 1×10^6 data collection iterations. The different runs for the same K were then averaged using the software CLUMPP (Jakobsson & Rosenberg, 2007). The most likely K was selected calculating an empirical statistic, ΔK, based on the second derivative of the likelihood of K (Evanno *et al.*, 2005). Including a reliable genetic structure is crucial to avoid false positives in GWAS. Consequently, the genetic structure obtained through Bayesian clustering was evaluated with an independent approach. For this purpose, we conducted a Principal Coordinates Analysis (PCoA) on a matrix of pair-wise genetic distances (G'_{ST}) between populations, and the first two components were plotted to summarise relations between populations. Both the calculation of genetic distances and PCoA were carried out using the program GenAlEx (Peakall & Smouse, 2006). Finally, the genetic clustering obtained with the two approaches was compared with evidence emerging from a previous study performed on a larger number of populations (Ruiz Daniels *et al.*, 2018).

GWAS

A GWAS was performed to test for the association between genotypes at single locus and each phenotypic trait (i.e. H_{UAV}, crown area, canopy temperature and vegetation indices). The residuals obtained through ANOVA (as phenotypes) were associated with genotypes at 235 SNPs out of the original 294 ones. These 235 SNPs were those showing high quality genotypes and a frequency of the minor allele >0.05. A total of 375 individuals, for which both genotypic and phenotypic data were available, were used for this analysis.

A mixed-effects linear model (MLM) was fitted independently for all pair combinations of SNPs and phenotypic traits following the procedure of Yu *et al.* (2006) implemented in Tassel

5.0 (Bradbury *et al.*, 2007). The probability of membership of each individual to each genetic cluster detected from clustering analysis (see "Results" section) was included to avoid false positive associations related to large-scale genetic structure. Moreover, the kinship matrix was also included in the model to control for finest genetic structure (Yu *et al.,* 2006). After the MLMs were fitted, a correction for multiple testing was performed on the *p*-values using the false discovery rate method (Storey & Tibshiriani, 2003) implemented in R qvalue package (Storey *et al.,* 2019).

SNP loci for which significant associations emerged with phenotypic traits were annotated from homology with other plant species as follows. First, the 200 bp (base pairs) sequence in which the SNPs where originally identified (Pinosio et *al.,* 2014) was used as query in the blastn tool. The best target sequence was subsequently used as query in blastx. The best match from blastx was retrieved to get the annotation of the SNPs. The result of blastn was also used to identify the effect of polymorphisms on proteins. To this purpose, the 200 bp sequence was aligned with the sequence of the best match derived from blastn. The sequence derived from this alignment was then translated to protein based on the Genbank information. The software DnaSP 6 (Rozas *et al.*, 2016) was employed to check if the SNPs were located in a non-coding or coding region and, in the latter case, the result change in the amino acids sequence (synonymous/non synonymous).

Multi-locus association tests

GWAS tests for significant contributions of a single locus to trait variation, but it does not provide an estimate for the total number of loci contributing to it. Alternatively, the number of SNPs explaining trait variation and the overall phenotypic variance explained by combinations of SNPs were estimated using the multi-locus mixed model (MLMM) approach proposed by Segura *et al.* (2012). This approach consists in a stepwise mixed-model regression with forward inclusion of SNPs markers. Prior to the stepwise analysis, a null model is calculated by incorporating all the possible SNPs, the kinship matrix and the genetic structure. The variance is then partitioned into genetic variance (i.e. explained by the kinship matrix, the population structure and the SNPs) and residual variance. At each step of the stepwise regression, the SNP with the lowest *p-value* is incorporated in the model, and the *p-values* of the other SNPs are re-calculated, as well as the explained genetic and residual variance. The stepwise regression is terminated when the explained genetic variance reaches a value proximal to that of the null model (i.e. incorporating all the SNPs). We used the extended Bayesian Information Criterion (eBIC; Chen & Chen, 2008) to select the model that best fitted our data among the different steps of the

MLMM. The SNPs included as cofactors and the percentage of variance explained (PVE) were recorded for the best fitting model. MLMM was performed for each phenotypic trait analysed in the GWAS.

RESULTS

UAV-derived individual crown delineation

A total of 736 trees (92%) were successfully identified through automatic canopy segmentation, while 72 trees were not recognized by the segmentation algorithm and were manually identified. There were also 13 cases of false tree identification. Altogether, 806 georeferenced crown shapes (discarding dead trees) were obtained (Fig. 2).

Canopy segmentation and PCA analysis

Based on the detected treetops and tree crowns, individual records of H_{UAV}, crown area, vegetation indices and canopy temperature were obtained for the 806 individuals of the common garden (Fig. S4). Mean H_{UAV} and crown area among trees were 5.80 ± 1.04 m and 6.52 ± 1.99 m^2 respectively (mean ± standard deviation). H_{UAV} correlated with ground-based height measurements at tree-level (r=0.85, p<0.001), indicating a reliable image-derived estimation of tree height. Crown area correlated with ground-measured DBH (r=0.70, p<0.001) (Fig. S5).

The PCA loadings summarised the relationships among vegetation indices, canopy temperature H_{UAV} and crown area (Fig. 3). Simple correlations between traits are reported in Table S2. Indices related to leaf area (i.e. NDVI, RDVI and OSAVI) grouped together in the plot of loadings, and they were unrelated to TCARI/OSAVI (informative of chlorophyll content). MCARI and TCARI (indicative of both leaf area and chlorophyll content) and EVI (indicative of leaf area) were grouped in between the other leaf area indices (NDVI, OSAVI and RDVI) and TCARI/OSAVI (Fig. 3; Table S2). WBI (informative of water content), ARI2 (informative of anthocyanin content) and CRI2 (informative of carotenoid content) were less explained by the first two PCA dimensions and were negatively associated with TCARI/OSAVI. Canopy temperature was negatively associated with H_{UAV} and crown area, which, in turn, were positively related to indices describing leaf area (Table S2). However, canopy temperature, H_{UAV} and crown area were poorly represented in the first two axes of the PCA (Fig. 3).

Neutral genetic data

The Bayesian clustering performed on SSRs genotypes indicated that the most probable number of clusters was two (Fig. S6). In this scenario, Western Mediterranean populations comprising

individuals from the Iberian Peninsula and the Balearic islands were separated from Central-Eastern Mediterranean ones (i.e. Greek, Italian and Tunisian) (Fig. 4). One Tunisian population showed a high degree of admixture. Possible clustering also emerged at K=4 or K=7, but with a lower probability than the K=2 scenario (Fig. S6). In addition, considering four or seven genetic clusters did not result in clear geographical distinction between genetic groups. The scenario with two main genetic clusters was confirmed by PCoA performed on a matrix of populations' genetic distances (Fig. S7). The first two axes of the PCoA identified a clear clustering in the two abovementioned groups. The scenario with two main genetic clusters also agreed with previously reported evidences (Ruiz Daniels *et al.*, 2018). Probabilities of individual assignment to the two clusters at K=2 were therefore used for correcting genotypic-phenotypic associations and avoid false positives.

GWAS

GWAS revealed twelve significant associations between SNPs and growth traits (i.e. H_{UAV} and crown area) with a p-value$< 10^{-3}$ (Table 2). These associations were significant also after correction for multiple testing (q-value< 0.05). H_{UAV} was associated with ten SNPs, which individually explained between 5% and 10% of phenotypic variance. Two SNPs were associated with crown area, each explaining *ca*. 5% of phenotypic variance.

Sixteen significant associations emerged between SNPs and vegetation indices (Table 2). Among those, seven associations occurred with indices related to leaf area, five with WBI (indicative of leaf water content), two with CRI2 (indicative of carotenoid content) and one with ARI2 (indicative of anthocyanin content). Most associations were also significant (q-value$<$ 0.05) or marginally significant (q-value$<$ 0.1) after correction for multiple testing, with the exception of the SNP associated with ARI2 (q-value$=$ 0.14). These associations explained individually a small proportion of phenotypic variance (range 3.5% - 5.5%). No associations emerged with TCARI/OSAVI (indicative of chlorophyll content). Finally, one SNP was associated with canopy temperature, explaining *ca.* 4% of phenotypic variance. However, this association was not significant after correction for multiple testing (q-value$=$ 0.11).

Some SNPs were associated with more than one trait (Table 2), as follows: SNP 91 was associated with EVI (indicative of leaf area) and WBI (indicative of leaf water content), SNP 108 was associated with H_{UAV}, EVI and WBI, SNP 133 was associated with H_{UAV} and WBI, SNP 241 was associated with EVI and MCARI, and, finally, SNP 273 and SNP 350 were both associated with H_{UAV} and EVI.

406 The majority of SNPs for which significant associations emerged resulted annotated in
407 known genes (Table 3; Table S3). For some of the annotated markers, it was possible to retrieve
408 the molecular and biological functions of the homologous protein. In most cases, however, SNPs
409 were located in non-coding regions or resulted in synonymous polymorphisms (Table S3).
410
411 **Multi-locus association test**
412 The multi-locus association test performed through MLMM revealed that combinations of the
413 235 SNPs explained a relatively high phenotypic variance of several traits (Table 4). Some, but
414 not all, of the SNPs identified by the single-locus association test were present in the multi-locus
415 association models. Notably, a combination of ten and nine SNPs explained 18% and 20% of
416 phenotypic variance of EVI (related to leaf area) and WBI (related to leaf water content)
417 respectively. In the case of H_{UAV}, 16% of phenotypic variance was explained by a combination
418 of only two SNPs, both identified also by GWAS. On the other hand, the best MLMM model
419 explained a low percentage of phenotypic variance for several traits including NDVI, RDVI,
420 OSAVI and those indices related to photosynthetic pigments. Finally, a combination of nine
421 SNPs explained 16% of phenotypic variance in canopy temperature.
422
423 **DISCUSSION**
424 **Canopy segmentation and recovery of phenotypic information**
425 Our study is, to the best of our knowledge, the first to apply automatic crown identification in a
426 common garden of a forest species. This approach was highly effective to identify single trees
427 and to provide tree-level estimations of growth-related traits. Although ground-based and UAV
428 data were collected over different years, the inter-tree variability in height estimated through
429 RGB-derived imagery agreed with ground-based measurements. Similarly, crown area was a
430 good indicator of stem diameter, as reported elsewhere for other species (Lockhart *et al.*, 2005;
431 Filipescu *et al.*, 2012), including *Pinus* spp. (Pretzsch et *al.,* 2015).
432 Developing cost and time-effective phenotyping approaches is fundamental for
433 characterizing the genetic basis of phenotypic differentiation (Lobos *et al.*, 2017). Growth
434 measurements can be obtained through ground-based approaches, but they require a significant
435 amount of fieldwork, especially in the case of adult individuals. On the other hand, our procedure
436 retrieved this information through a single flight and few hours of computation. In this regard,
437 crown identification based on RGB-derived dense point clouds is being increasingly used for the
438 estimation of growth traits in natural forests and fruit orchards (Wallace *et al.*, 2016; Weiss &
439 Baret, 2017). Several studies have shown that this approach – based on inexpensive devices –

can provide accurate estimations comparable to those obtained with more expensive and complex technologies such as light detection and ranging (LiDAR) (Zarco-Tejada *et al.*, 2014; Wallace *et al.*, 2016). In this regard, the continuous development and optimization of algorithms for automatic crown identification can facilitate the implementation of this methodology to forest genetic trials. Another advantage is that it allows the estimation of tree-level values of vegetation indices, which are surrogates of meaningful morpho-physiological traits (Roberts *et al.*, 2016; Santini *et al.*, 2019b). Traits such as leaf area or canopy transpiration can hardly be retrieved for a significant number of adult trees through ground-based phenotyping techniques. Hence, the approach developed in this study could allow for easy routine analysis of barely investigated phenotypic traits so far.

In a previous study performed in the same common garden, vegetation indices and canopy temperature derived at plot level were used to investigate inter-population differentiation in vegetation characteristics (Santini *et al.*, 2019a). The present study provided tree-level information which is suited for genotype-phenotype association studies and the assessment of relationships among phenotypic traits and their trade-offs at individual level. In this regard, the PCA analysis revealed strong relationships between indices related to leaf area, canopy temperature and aboveground growth traits. Trees having high leaf area and reduced canopy temperature (indicative of high transpiration; González-Dugo *et al.*, 2013) showed high growth, indicated by large height and crown area. These results confirm the strong dependence of growth to variation in photosynthetic surface and total transpiration in *P. halepensis* under drought conditions (Voltas *et al.*, 2008; Santini *et al.*, 2019a). On the other hand, variation in leaf biochemistry seemed not to affect tree growth to the same degree (Santini *et al.*, 2019a).

Association genetics of growth traits

Despite the low number of markers considered, the results of this study showed the effectiveness of combining genotypic information with UAV imagery to characterise the genomic basis of phenotypic differentiation in forest species. Indeed, we identified relevant genotypic-phenotypic associations for several traits related to growth and vegetation indices. Such associations provide a first insight into adaptive genomic variation in *P. halepensis*, although an intrinsic limitation of our approach is the lack of strong evidence on how variation at the cellular and tissue levels is coordinated with changes at the whole-organism level. Anyhow, the detected loci deserve further attention in relation to their potential adaptive role for this and other closely related pines.

GWAS identified 12 SNPs significantly associated with UAV-derived crown area and tree height, explaining a proportion of phenotypic variance of 5-10%. Since linkage

disequilibrium decays rapidly in conifers (De La Torre *et al.*, 2014; Plomion *et al.*, 2016), these SNPs are likely quantitative trait nucleotides (i.e. single polymorphisms influencing the phenotype) or are located in close proximity to the causative polymorphisms. Our results indicate that these SNPs do not produce protein change or alteration, being probably involved in changes of gene expression. Most of the relevant SNPs were annotated with known proteins, even though a correct annotation was not feasible in some cases. This may be partially due to the relatively scarce information on conifer genomes that is available to date, which emphasises the need of further studies to characterise tree genomes (De La Torre *et al.*, 2014).

Growth-associated SNPs that could be correctly annotated included genomic regions encoding for a disparate number of proteins. In particular, differences in crown area were associated with a gene marker (SNP 9) encoding for a heat stress transcription factor, which is part of a family of proteins involved in tolerance and response to heat stress (Scharf *et al.*, 2012). In the case of tree height, significant associations were found with genes encoding for (1) calcium-dependent kinase (SNP 18 and SNP 159), which influences plant responses to endogenous and environmental cues (Romeis, 2001), (2) GDP-mannose pyrophosporylase (SNP 340) involved in ascorbic acid metabolism (Keller *et al.*, 1999), (3) cellulose synthase (SNP 273), which influences cell wall synthesis and organization (Richmond & Somerville, 2000), and (4) polygalacturonase (SNP 133 and SNP 206) involved in tissue development and stomatal dynamics (Rui *et al.*, 2017).

These results suggest that the genetic control of complex traits such as growth is associated with a large number of interacting genes with very diverse functions. However, the multi-locus association analysis revealed that a combination of only two SNPs explained most phenotypic variation (16%) in tree height. This outcome could be partially due to the inability of the MLMM approach to identify loci with small effects in the presence of loci with large effects on the phenotype. On the other hand, the result of the MLMM may also indicate that the effects of single locus of tree height identified by GWAS are largely overlapping. Finally, it must be noted that differences in the number of SNPs detected through each approach could originate, at least in part, from different statistical treatments of missing data.

Association genetics of vegetation properties

GWAS also revealed a number of SNPs significantly associated with phenotypic variation in vegetation indices. These associations were inconsistent across indices related to leaf area, however. Some indices (i.e. NDVI, RDVI and OSAVI) are known to saturate at high leaf areas, being potentially unable to intercept fine inter-individual differences in this trait (Roberts *et al.*,

2016). In contrast, several marker-trait associations emerged with EVI, which is less sensitive to saturation. Two of the identified SNPs related to EVI were annotated to known proteins. SNP 91 encodes for a clavata-like receptor kinase, which is involved in tissue developing and meristem differentiation (Clark *et al.*, 2001). SNP 273 is found in a gene encoding for a cellulose synthase, which is part of a family of genes with a crucial role in cell wall synthesis and organization, in turn influencing tissue growth and development (Richmond & Somerville, 2000). Interestingly, proteins of the cellulose synthase family are associated with the control of leaf size in *Arabidopsis* (Horiguchi *et al.*, 2006) and maize (Li *et al.*, 2018). Another gene marker (SNP 151) was associated with TCARI, an index partly related to leaf area. SNP 151 is found in a gene encoding for PEX11C protein involved in peroxisome metabolism. Among other functions, proteins of the PEX family have been shown to regulate photomorphogenesis through a light-mediated pattern of induction of peroxisome proliferation, with potential implications on leaf development (Hu *et al.,* 2002; Kaur *at al.*, 2013). Interestingly, this SNP was reported to be under selection in *P. halepensis* and associated with variation in summer precipitation at the origin of natural populations (Ruiz Daniels *et al*. 2018). In other Mediterranean pines (*P. pinaster* and *P. pinea*), the expression of a homologous gene was induced by drought stress (Perdiguero et *al.,* 2013).

Although TCARI is also sensitive to changes in leaf chlorophyll content, no significant marker-trait associations emerged with the index TCARI/OSAVI, which is specifically related to chlorophyll content (Zarco-Tejada *et al.*, 2004). This may be due to weak genetic differentiation in photosynthetic capacity of *P. halepensis* (Santini *et al.*, 2019a). On the other hand, four SNPs were associated with carotenoid or anthocyanin content, and two of them were annotated with homologous proteins described in other species. In particular, SNP 258 is located in a gene encoding for a PIN2 protein involved in auxin transport and auxin-mediated signalling, while the protein associated with SNP 67 is involved in pathogen resistance. Anthocyanins and carotenoids are involved in a myriad of plant processes related to stress response (Havaux, 2014; Van der Ende & El-Esawe**,** 2014). Interestingly, some studies have reported that carotenoid accumulation is, in some cases, mediated by auxin balance (Du *et al.*, 2012; Su *et al.*, 2015). In this regard, SNP 258 was also identified in a recent study as candidate gene involved in adaptation to local climate (Ruiz Daniels *et al.*, 2019).

The WBI index, indicative of leaf water content, was associated with five SNPs. Two of them are located in genes of known functions. SNP 133 is found in a gene encoding the polygalacturonase isoform X3 protein, which has been widely studied in the model plant *Arabidopsis*. This protein influences tissue development and is involved in stomatal dynamics

of mature leaves, which might explain the association found with leaf water content (Rui *et al.*, 2017). SNP 91 encodes for a clavata-like receptor kinase and it was associated also with leaf area indices (Clark *et al.*, 2001).

However, the identified SNPs had small effects on phenotypes, explaining individually less than 6% of the variation in vegetation indices or canopy temperature. Genetic variation in single markers usually accounts for a small proportion of phenotypic variance in forest species (Eckert *et al.*, 2009; Prunier *et al.*, 2013; Baison *et al.*, 2019). Indeed, many adaptive traits are likely under polygenic control, suggesting that many loci with minor effect might modify their expression (White *et al.,* 2007; Savolainen *et al.,* 2007). Moreover, the single contribution of many other genes to complex traits such as leaf architecture or water use may be too small to be detected by GWAS (Resende *et al.*, 2017). In this regard, the multi-locus association test showed that combinations of several markers are able to explain a relatively large phenotypic variation in leaf area and transpiration. This result suggests that association analyses carried out considering a larger number of markers simultaneously can capture a sizeable variation of complex traits in forest species (Savolainen *et al.*, 2007; Grattapaglia *et al.*, 2009).

Even considering a polygenic model, however, a large part of phenotypic variation both in vegetation indices and in growth traits remained unexplained. We speculate that this may be related to inaccuracies in UAV-derived records and, also, to low marker coverage, which stresses the importance of allowing for the largest possible number of markers across the genome (Resende *et al.*, 2017). Nevertheless, it is worth noting that a high phenotypic variance is seldom explained by genetic variation even in organisms for which dense, genome-wide sets of markers are available (Manolio *et al.,* 2009; Bourrat *et al.,* 2017). This issue, known as "missing hereditability", suggests that phenotypic variation may be broadly influenced by largely unknown underlying factors such as epistatic interactions or epigenetic modifications of gene expression (Manolio *et al.,* 2009; Trerotola *et al.,* 2015; Bourrat *et al.,* 2017).

Pleiotropy

Some SNPs showed significant associations with multiple correlated traits, a possible indication of pleiotropic effects with genes influencing different traits simultaneously (Wu *et al.,* 2000). However, these SNPs likely concur to determine only one trait, showing multiple associations because of correlated traits. For example, three SNPs (108, 273 and 350) were associated with both tree height and some indices related to leaf area. Indeed, aerial growth and photosynthetic surface described by vegetation indices were associated in *P. halepensis*, both at tree (this study) and population level (Santini *et al.*, 2019a). On the other hand, SNPs 133 and 350 were

associated with both tree height and WBI, traits that were weakly correlated among individuals. Such associations involving more than one trait could be indicative of true pleiotropy, with gene markers influencing both leaf water content and tree growth concurrently.

Conclusions

In this study, we developed a workflow for (semi-) automatic phenotyping of trees growing in a common garden. In the era of genomics, retrieving meaningful phenotypic information with time and cost effective tools has the potential to boost the characterisation of the genetic basis of fundamental evolutionary processes involved in phenotypic differentiation of non-model organisms (Großkinsky *et al.*, 2015). Although the scope of our inferences could be undoubtedly broadened by using a significantly larger number of markers and individuals, this study highlights the type of information that can be obtained through high-throughput phenotyping approaches. Indeed, our results provide an insight on the molecular processes controlling phenotypic differentiation in a widespread conifer, underlining the potential of widely available new technologies to fill the gap between genetic variability and individual phenotypes (Houle *et al.*, 2010).

Acknowledgements

This work was supported by the Spanish Government, grant number RTI2018-094691-B-C31 (MCIU/AEI/FEDER, EU). We thank Ricardo Alía (CIFOR-INIA) for providing the genotypic data. We thank three anonymous reviewers whose insightful comments and suggestions helped to improve the manuscript

Author contributions

JV and FS conceived and designed the research. FS, SCK, VRD, DG and JV collected the data; FS and SMG analysed the data; FS and JV wrote the manuscript, with contributions from SCK, JLA, VRD, SMG and DG.

Data Availability

Individual genotypic data are available at Dryad repository [https://datadryad.org/stash/dataset/doi:10.5061/dryad.pt3b974]. Phenotypic data will be made available at a Mendeley repository.

610 **Conflict of interest**

611 None

612 **REFERENCES**

613 **Baison J, Vidalis A, Zhou L, Chen Z, Li Z, Sillanpä MJ, Bernhardsson C, Scofield D,**
614 **Forsberg N, Grahn T et al. 2019.** Genome-wide association study identified novel candidate
615 loci affecting wood formation in Norway spruce. *The Plant Journal* **100:** 83-100.

616

617 **Bourrat P, Lu Q, Jablonka E. 2017.** Why the missing heritability might not be in the DNA.
618 *BioEssays* **39:** 1700067.

619

620 **Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007.**
621 TASSEL: software for association mapping of complex traits in diverse samples.
622 *Bioinformatics* **23:** 2633-2635.

623

624 **Chen J, Chen Z. 2008.** Extended Bayesian information criteria for model selection with large
625 model spaces. *Biometrika* **95:** 759-771.

626

627 **Clark SE. 2001.** Cell signalling at the shoot meristem. *Nature Reviews Molecular Cell Biology*
628 **2:** 276-284.

629

630 **Climent J, Prada MA, Calama R, Chambel MR, de Ron DS, Alía R. 2008.** To grow or to
631 seed: ecotypic variation in reproductive allocation and cone production by young female
632 Aleppo pine (*Pinus halepensis* , Pinaceae). *American Journal of Botany* **95:** 833–842.

633

634 **Daughtry C. 2000.** Estimating corn leaf chlorophyll concentration from leaf and canopy
635 reflectance. *Remote Sensing of Environment* **74:** 229–239.

636

637 **De La Torre AR, Birol I, Bousquet J, Ingvarsson PK, Jansson S, Jones SJM, Keeling CI,**
638 **MacKay J, Nilsson O, Ritland K et al. 2014.** Insights into conifer giga-genomes. *Plant*
639 *Physiology* **166:** 1724–1732.

Du H, Wu N, Fu J, Wang S, Li X, Xiao J, Xiong L. 2012. A GH3 family member, OsGH3-2, modulates auxin and abscisic acid levels and differentially affects drought and cold tolerance in rice. *Journal of Experimental Botany* **63:** 6467-6480.

Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutovsky KV, St. Clair JB, Neale D B. 2009. Association genetics of coastal Douglas fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-hardiness related traits. *Genetics* **182:** 1289-1302.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14:** 2611–2620.

Filipescu CN, Groot A, MacIsaac DA, Cruickshank MG, Stewart JD. 2012. Prediction of diameter using height and crown attributes: a case study. *Western Journal of Applied Forestry* **27:** 30–35.

Gitelson AA, Merzlyak MN, Chivkunova OB. 2001. Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochemistry and Photobiology* **74:** 38–45.

Gitelson AA, Zur Y, Chivkunova OB, Merzlyak MN. 2002. Assessing carotenoid content in plant leaves with reflectance spectroscopy. *Photochemistry and Photobiology* **75:** 272–281.

González-Dugo V, Zarco-Tejada P, Nicolás E, Nortes PA, Alarcón JJ, Intrigliolo DS, Fereres E. 2013. Using high resolution UAV thermal imagery to assess the variability in the water status of five fruit tree species within a commercial orchard. *Precision Agriculture* **14:** 660–678.

668 **Grattapaglia D, Plomion C, Kirst M, Sederoff RR. 2009.** Genomics of growth traits in forest
669 trees. *Current Opinion in Plant Biology* **12:** 148–156.

670

671 **Grattapaglia D, Resende MDV. 2011.** Genomic selection in forest tree breeding. *Tree*
672 *Genetics & Genomes* **7:** 241–255.

673

674 **Großkinsky DK, Svensgaard J, Christensen S, Roitsch T. 2015.** Plant phenomics and the
675 need for physiological phenotyping across scales to narrow the genotype-to-phenotype
676 knowledge gap. *Journal of Experimental Botany* **66:** 5429–5440.

677

678 **Havaux M. 2014.** Carotenoid oxidation products as stress signals in plants. *The Plant Journal*
679 **79:** 597-606.

680

681 **Haboudane D, Miller JR, Tremblay N, Zarco-Tejada PJ, Dextraze L. 2002.** Integrated
682 narrow-band vegetation indices for prediction of crop chlorophyll content for application to
683 precision agriculture. *Remote Sensing of Environment* **81:** 416–426.

684

685 **Hardy O, Vekemans X. 2015.** *SPAGeDi 1.5. a program for spatial pattern analysis of genetic*
686 *diversity*. Brussells, Belgium: Eds. Université Libre de Bruxelles.

687

688 **Horiguchi G, Ferjani A, Fujikura U, Tsukaya H. 2006.** Coordination of cell proliferation
689 and cell expansion in the control of leaf size in Arabidopsis thaliana. *Journal of Plant Research*
690 **119:** 37–42.

691

692 **Houle D, Govindaraju DR, Omholt S. 2010.** Phenomics: the next challenge. *Nature Reviews*
693 *Genetics* **11:** 855–866.

694

695

Hu J, Aguirre M, Peto C, Alonso J, Ecker J, Chory J. 2002. A role for peroxisomes in photomorphogenesis and development of Arabidopsis. *Science* **297:** 405-409.

Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment* **83:** 195–213.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23:** 1801–1806.

Jaramillo-Correa JP, Prunier J, Vázquez-Lobo A, Keller SR, Moreno-Letelier A. 2015. Molecular signatures of adaptation and selection in forest trees. *Advances in Botanical Research* **74:** 265–306.

Kaur N, Li J, Hu J. 2013. Peroxisomes and photomorphogenesis. In Luis A, Rio D, eds. *Peroxisomes and their key role in cellular signaling and metabolism*. Dordrecht, Netherlands: Eds. Springer, 195-211.

Keller R, Renz FSA, Kossmann J. 1999. Antisense inhibition of the GDP-mannose pyrophosphorylase reduces the ascorbate content in transgenic plants leading to developmental changes during senescence. *The Plant Journal* **19:** 131–141.

Khan MA, Korban SS. (2012) Association mapping in forest trees and fruit crops. *Journal of Experimental Botany* **63:** 4045-4060.

Klein T, Di Matteo G, Rotenberg E, Cohen S, Yakir D. 2013. Differential ecophysiological response of a major Mediterranean pine species across a climatic gradient. *Tree Physiology* **33:** 26–36.

Li W, Yang Z, Yao J, Li J, Song W, Yang X. 2018. Cellulose synthase-like D1 controls organ size in maize. *BMC Plant Biology* **18:** 239.

Lobos GA, Camargo AV, del Pozo A, Araus JL, Ortiz R, Doonan JH. 2017. Editorial: plant phenotyping and phenomics for plant preeding. *Frontiers in Plant Science* **8:** 2181

Lockhart BR, Weih RC, Smith KM. 2005. Crown radius and diameter at breast height relationships for six bottomland hardwood species. *Journal of the Arkansas Academy of Science* **59:** 110-115

Loiselle BA, Sork VL, Nason J, Graham C. 1995. Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany* **82:** 1420–1425.

Ludovisi R, Tauro F, Salvati R, Khoury S, Mugnozza Scarascia G, Harfouche A. 2017. UAV-based thermal imaging for high-throughput field phenotyping of black poplar response to drought. *Frontiers in Plant Science* **8:** 1681

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ *et al.* 2009. Finding the missing heritability of complex diseases. *Nature* **461:** 747-753.

McKown AD, Klápště J, Guy RD, Geraldes A, Porth I, Hannemann J, Friedmann M, Muchero W, Tuskan GA, Ehlting J et al. 2014. Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. *New Phytologist* **203:** 535–553.

McGaughey RJ. 2012. *FUSION/LDV: Software for LiDAR data analysis and visualization—V3. 10*. Portland, OR, USA: Eds. USDA Forest Service, Pacific Northwest Research Station

753 **Meyer F, Beucher S. 1990.** Morphological segmentation. *Journal of Visual Communication*
754 *and Image Representation* **1:** 21-46.

755

756 **Nevalainen O, Honkavaara E, Tuominen S, Viljanen N, Hakala T, Yu X *et al.* 2017.**
757 Individual tree detection and classification with UAV-based photogrammetric point clouds and
758 hyperspectral imaging. *Remote Sensing* **9:** 185.

759

760 **Peakall R, Smouse PE. 2006.** Genalex 6: genetic analysis in Excel. Population genetic
761 software for teaching and research. *Molecular Ecology Notes* **6:** 288–295.

762

763 **Peñuelas J, Filella I, Biel C, Serrano L, Save R. 1993.** The reflectance at the 950–970 nm
764 region as an indicator of plant water status. *International Journal of Remote Sensing* **14:** 1887–
765 1905.

766

767 **Perdiguero P, del Carmen Barbero M, Cervera MT, Collada C. 2013.** Molecular response
768 to water stress in two contrasting Mediterranean pines (*Pinus pinaster* and *Pinus pinea*). *Plant*
769 *Physiology and Biochemistry* **67:** 199-208.

770

771 **Pinosio S, González-Martínez SC, Bagnoli F, Cattonaro F, Grivet D, Marroni F, Lorenzo**
772 **Z, Pausas JG, Verdú M, Vendramin GG. 2014.** First insights into the transcriptome and
773 development of new genomic tools of a widespread circum-Mediterranean tree species, *Pinus*
774 *halepensis* Mill. *Molecular Ecology Resources* **14:** 846–856.

775

776 **Plomion C, Bartholomé J, Lesur I, Boury C, Rodríguez-Quilón I, Lagraulet H,**
777 **Ehrenmann F, Bouffier L, Gion JM, Grivet D et al. 2016.** High-density SNP assay
778 development for genetic analysis in maritime pine (*Pinus pinaster*). *Molecular Ecology*
779 *Resources* **16:** 574–587.

780

Plowright A. 2018. *Foresttools: analyzing remotely sensed forest data.* R Package Version 0.2.0.

Popescu SC, Wynne RH. 2004. Seeing the trees in the forest. *Photogrammetric Engineering & Remote Sensing* **70:** 589-604.

Pretzsch H, Biber P, Uhl E, Dahlhausen J, Rötzer T, Caldentey J *et al.* 2015. Crown size and growing space requirement of common tree species in urban centres, parks, and forests. *Urban Forestry & Urban Greening* **14:** 466-479.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155:** 945-959.

Prunier J, Pelgas B, Gagnon F, Desponts M, Isabel N, Beaulieu J, Bousquet J. 2013. The genomic architecture and association genetics of adaptive characters using a candidate SNP approach in boreal black spruce. *BMC Genomics* **14:** 368.

QGIS Development Team. 2019. *QGIS Geographic Information System. Open Source Geospatial Foundation Project.* [WWW document] URL http://qgis.osgeo.org/ [accessed 20 January 2020].

R Core Team (2019). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. [WWW document] URL https://www.R-project.org/ [accessed 20 January 2020].

Resende RT, Resende MDV, Silva FF, Azevedo CF, Takahashi EK, Silva-Junior OB, Grattapaglia D. 2017. Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in *Eucalyptus*. *New Phytologist* **213:** 1287–1300.

Richmond TA, Somerville CR. 2000. The cellulose synthase superfamily. *Plant Physiology* **124:** 495-498.

Roberts DA, Roth KL, Perroy RL. 2016. Hyperspectral vegetation indices. In: Huete A., Lyon JG, Thenkabail PS, eds. *Hyperspectral Remote Sensing of Vegetation*. Boca Raton, FL, USA: CRC Press, 309–328.

Romeis T. 2001. Calcium-dependent protein kinases play an essential role in a plant defence response. *The EMBO Journal* **20:** 5556–5567.

Rondeaux G, Steven M, Baret F. 1996. Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment* **55:** 95–107.

Roujean J-L, Breon F-M. 1995. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment* **51:** 375–384.

Rouse J, Haas RH, Schell JA, Deering DW. 1974. Monitoring vegetation systems in the Great Plains with ERTS. *NASA Special Publication* **351:** 309

Roussel JR, Auty D. 2018. *lidR: Airborne LiDAR data manipulation and visualization for forestry applications.* R package version 1.0.0.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico, S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large datasets. *Molecular Biology and Evolution* **34:** 3299-3302.

Rui Y, Xiao C, Yi H, Kandemir B, Wang JZ, Puri VM, Anderson CT. 2017. Polygalacturonase involved in expansion3 functions in seedling development, rosette growth, and stomatal dynamics in *Arabidopsis thaliana*. *The Plant Cell* **29:** 2413–2432.

Ruiz Daniels R, Taylor RS, Serra-Varela MJ, Vendramin GG, González-Martínez SC, Grivet D. 2018. Inferring selection in instances of long-range colonization: The Aleppo pine (*Pinus halepensis*) in the Mediterranean Basin. *Molecular Ecology* **27:** 3331–3345.

Ruiz Daniels R, Taylor RS, González-Martínez SC, Vendramin GG, Fady B, Oddou-Muratorio S, Piotti A, Simioni G, Grivet D, Beaumont MA. 2019. Looking for local adaptation: convergent microevolution in Aleppo Pine (*Pinus halepensis*). *Genes* **10:** 673.

Santini F, Kefauver SC, Resco de Dios V, Araus JL, Voltas J. 2019a. Using unmanned aerial vehicle-based multispectral, RGB and thermal imagery for phenotyping of forest genetic trials: a case study in *Pinus halepensis*. *Annals of Applied Biology* **174:** 262–276.

Santini F, Serrano L, Kefauver SC, Abdullah-Al M, Aguilera M, Sin E, Voltas J. 2019b. Morpho-physiological variability of Pinus nigra populations reveals climate-driven local adaptation but weak water use differentiation. *Environmental and Experimental Botany* **166:** 103828.

Savolainen O, Pyhäjärvi T, Knürr T. 2007. Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution and Systematics* **38:** 595–619.

Scharf KD, Berberich T, Ebersberger I, Nover L. 2012. The plant heat stress transcription factor (Hsf) family: structure, function and evolution. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **1819:** 104-119.

Schiller G, Atzmon N. 2009. Performance of Aleppo pine (*Pinus halepensis*) provenances grown at the edge of the Negev desert: A review. *Journal of Arid Environments* **73:** 1051–1057.

Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics* **44:** 825.

Serra-Varela MJ, Alía R, Daniels RR, Zimmermann NE, Gonzalo-Jiménez J, Grivet D. 2017. Assessing vulnerability of two Mediterranean conifers to support genetic conservation management in the face of climate change (L Brotons, Ed.). *Diversity and Distributions* **23:** 507–516.

Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* **100:** 9440-9445.

Storey JD, Bass AJ, Dabney A, Robinson D. 2019. *qvalue: Q-value estimation for false discovery rate control.* R package version 2.16.0.

Su L, Diretto G, Purgatto E, Danoun S, Zouine M, Li Z, Roustan JP, Bouzayen M, Giuliano G, Chervin C. 2015. Carotenoid accumulation during tomato fruit ripening is modulated by the auxin-ethylene balance. *BMC Plant Biology* **15:** 114.

Tognetti R, Michelozzi M, Giovannelli A. 1997. Geographical variation in water relations, hydraulic architecture and terpene composition of Aleppo pine seedlings from Italian provinces. *Tree Physiology* **17:** 241-250.

Trerotola M, Relli V, Simeone P, Alberti S. 2015. Epigenetic inheritance and the missing heritability. *Human Genomics* **9:** 17.

Van den Ende W, El-Esawe SK. 2014. Sucrose signaling pathways leading to fructan and anthocyanin accumulation: a dual function in abiotic and biotic stress responses?. *Environmental and Experimental Botany* **108:** 4-13.

de Villemereuil P, Gaggiotti OE, Mouterde M, Till-Bottraud I. 2016. Common garden experiments in the genomic era: new perspectives and opportunities. *Heredity* **116:** 249–254.

Voltas J, Chambel MR, Prada MA, Ferrio JP. 2008. Climate-related variability in carbon and oxygen stable isotopes among populations of Aleppo pine grown in common-garden tests. *Trees* **22:** 759–769.

Voltas J, Lucabaugh D, Chambel MR, Ferrio JP. 2015. Intraspecific variation in the use of water sources by the circum-Mediterranean conifer *Pinus halepensis*. *New Phytologist* **208:** 1031–1041.

Voltas J, Shestakova TA, Patsiou T, di Matteo G, Klein T. 2018. Ecotypic variation and stability in growth performance of the thermophilic conifer *Pinus halepensis* across the Mediterranean basin. *Forest Ecology and Management* **424:** 205–215.

Wallace L, Lucieer A, Malenovský Z, Turner D, Vopěnka P. 2016. Assessment of forest structure using two uav techniques: a comparison of airborne laser scanning and structure from motion (sfm) point clouds. *Forests* **7:** 62.

Weiss M, Baret F. 2017. Using 3D Point Clouds Derived from UAV RGB imagery to describe vineyard 3D macro-structure. *Remote Sensing* **9:** 111.

White TL, Adams WT, Neale DB. 2007. *Forest genetics.* Cambridge, MA, USA: Eds. CABI.

Wu R, Zeng ZB, McKeand SE, O Malley DM. 2000. The case for molecular mapping in forest tree breeding. *Plant Breeding Reviews* **19:** 41-68.

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38:** 203–208.

Zarco-Tejada PJ, Diaz-Varela R, Angileri V, Loudjani P. 2014. Tree height quantification using very high resolution imagery acquired from an unmanned aerial vehicle (UAV) and automatic 3D photo-reconstruction methods. *European Journal of Agronomy* **55:** 89–99.

Zarco-Tejada PJ, Miller JR, Morales A, Berjón A, Agüera J. 2004. Hyperspectral indices and model simulation for chlorophyll estimation in open-canopy tree crops. *Remote Sensing of Environment* **90:** 463–476.

Supporting Information

Table S1. Geographic origin of the 56 *Pinus halepensis* populations tested in this study.

Table S2. Correlation matrix between phenotypic traits.

Table S3. First step of the SNPs annotation.

Fig. S1. Aerial view of the common garden in July 2016.

Fig. S2. Geographic locations of the populations tested in the common garden.

Fig. S3. Schematized process for estimation of tree height and crown area, and for obtaining crown multispectral and thermal images.

Fig. S4. Boxplots of inter-tree variation in growth data, canopy temperature and vegetation indices.

Fig. S5. Correlations between ground measured and UAV-derived growth data.

Fig. S6. ΔK results from STRUCTURE analysis.

Fig. S7. Plot of first and second coordinates of PCoA performed on a matrix of genetic distances between populations.

Methods Supporting Information

Figure legends

Figure 1. Flowchart showing the different steps followed to carry out the study of association genetics in *P. halepensis* using RGB, multispectral and thermal imagery as phenotyping tools.

Figure 2. a) Digital terrein model and b) canopy elevation model obtained through structure-from-motion analysis of RGB images. The correctly identified treetops (red dots), the false positive tree tops (i.e. no trees identified as trees; light blue dots) and the false negative treetops (i.e. trees not identified) are plotted on the canopy elevation model together with the segmented crown shapes (red polygons). An area of the trial (black square) has been enlarged to show the difference between crown shaped before (c) and after (d) the manual correction of the mis-idenfied treetops.

Figure 3. Plot of the loadings of the first two PCA axes describing the relations between UAV-derived growth traits, vegetation indices and canopy temperature. NDVI, Normalized Difference Vegetation Index; OSAVI, Optimised Soil Adjusted Vegetation Index; RDVI, Re-normalized Difference Vegetation Index; EVI, Enhanced Vegetation Index; MCARI, Modified Chlorophyll Absorption Reflectance Index; TCARI, Transformed Chlorophyll Absorption Ratio Index; CRI2 Carotenoid Reflectance Index 2; ARI2, Anthocyanin Reflectance Index 2; WBI Water Band Index; T, canopy temperature; H_{UAV}, UAV-retrieved tree height.

Figure 4. a) Bar plot showing results of the assignment test with K=2 for the 375 genotyped individuals. Each individual is represented by a vertical line divided into two coloured segments representing the probabilities that the individual is assigned to the group comprising Western (in red) or Central-Eastern (in blue) Mediterranean populations. Population tags are reported in Table S1. b) Pie charts showing the percentage of assignment to the two groups in each population. The charts are plotted on the actual geographic coordinates of populations.

997 # Tables

Table 1. Vegetation indices (VIs) considered in this study. R indicates the reflectance in a single wavelength (in nm).

Index	Descriptor	Wavelengths	Formula	Reference
NDVI	Leaf area	Red, NIR	$(R_{840} - R_{670})/(R_{840} + R_{670})$	Rouse *et al.,* (1974)
OSAVI	Leaf area	Red, NIR	$(R_{840} - R_{670})/(R_{840} + R_{670} + 0.16)\times 1.16$	Rondeaux *et al.*, (1996)
RDVI	Leaf area	Red, NIR	$(R_{840} - R_{670})/(R_{840} + R_{670})^{\wedge 1/2}$	Roujean & Breon, (1995)
EVI	Leaf area	Blue, Red, NIR	$2.5\times(R_{840} - R_{670})/[(R_{840} + 6\times R_{670} - 7.5\times R_{450}) + 1]$	Huete *et al.*, (2002)
MCARI	Leaf chlorophyll content; leaf area	Green, Red, NIR	$[(R_{700} - R_{670})\text{-}0.2\times(R_{700} - R_{550})]\times(R_{700}/R_{670})$	Daughtry, (2000)
TCARI	Leaf chlorophyll content; leaf area	Green, Red, NIR	$3\times(R_{700} - R_{670})\text{-}0.2\times(R_{700} - R_{550})\times(R_{700}/R_{670})$	Haboudane *et al.*, 2002)
TCARI/OSAVI	Leaf chlorophyll content	Green, Red, NIR	-	Haboudane *et al.*,(2002)
ARI2	Anthocyanin content	Blue, NIR	$R_{840}\times(1/R_{550} - 1/R_{700})$	Gitelson *et al.,* (2001)
CRI2	Carotenoid content	Blue, NIR	$1/R_{550} - 1/R_{700}$	Gitelson *et al.,* (2002)
WBI	Water content	NIR	R_{900} / R_{950}	Peñuelas *et al.,* (1993)

Table 2. Results of the GWAS. The SNPs associated with traits with p-value < 10^{-3} are reported, together with the corrected q-value and the percentage of variance explained (PVE).

Trait	Indicator	Marker	q-value	PVE (%)
Leaf area				
	EVI	[A]SNP108	0.01	4.26
	EVI	[A]SNP241	0.01	3.75
	EVI	[B]SNP273	0.01	4.64
	EVI	[A]SNP350	0.01	4.43
	EVI	[A]SNP91	0.01	4.60
	MCARI	[A]SNP241	0.07	3.52
	TCARI	[A]SNP151	0.07	4.54
Photosynthetic pigments				
	ARI2	[B]SNP258	0.14	4.87
	CRI2	[A]SNP67	0.05	5.00
	CRI2	[B]SNP201	0.06	3.95
	CRI2	[B]SNP204	0.03	4.96
Water content				
	WBI	[A]SNP91	0.02	5.15
	WBI	[A]SNP108	0.02	4.07
	WBI	[A]SNP133	0.02	4.28
	WBI	[B]SNP265	0.02	4.46
	WBI	[A]SNP350	0.02	4.82
Canopy temperature				
	T	[B]SNP140	0.11	4.34
Growth traits				
	Crown Area	[B]SNP2	0.05	4.94
	Crown Area	[A]SNP9	0.05	5.38
	H_{UAV}	[A]SNP18	<0.001	7.39
	H_{UAV}	[A]SNP108	<0.001	6.64
	H_{UAV}	[A]SNP133	<0.001	6.43
	H_{UAV}	[A]SNP159	<0.001	6.85
	H_{UAV}	[B]SNP206	<0.001	5.70
	H_{UAV}	[A]SNP217	<0.001	5.08
	H_{UAV}	[B]SNP250	<0.001	5.50
	H_{UAV}	[B]SNP273	<0.001	5.52
	H_{UAV}	[B]SNP340	<0.001	5.70
	H_{UAV}	[A]SNP350	<0.001	10.03

[A]from re-sequences genes of loblolly pine; [B]from *P. halepensis* transcriptome

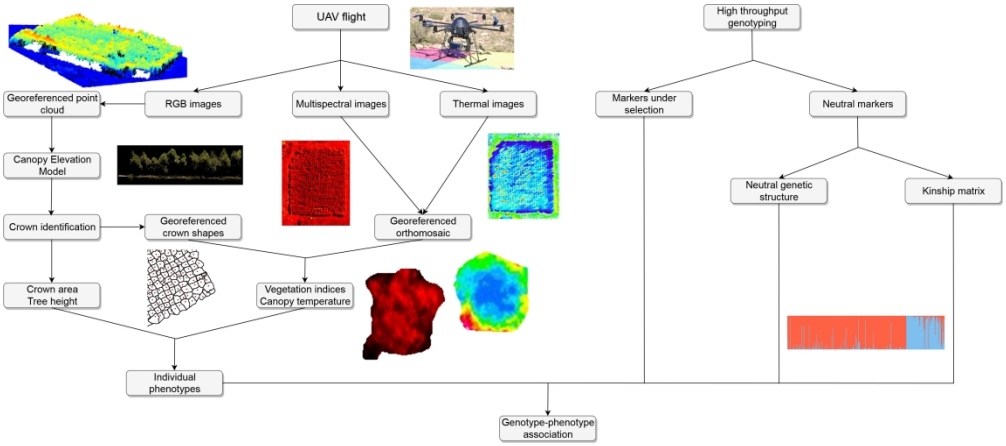
1018 **Table 3.** Annotation of the SNPs detected in GWAS.

SNP ID	Accession	E-values balstx	Annotation	Species	Molecular Function	Biological Function
SNP9	XP_006826240.1	2E-10	heat shock factor protein HSF24	*Amborella trichopoda*	DNA-binding transcription factor activity; RNA polymerase II cis-regulatory region; sequence-specific DNA	Cellular response to heat; regulation of transcription from RNA polymerase II promoter in response to stress
SNP18	ACJ09662.1	9E-45	putative calcium-dependent protein kinase	*Cupressus sempervirens*	Kinase; transferase	-
SNP67	XP_031397970.1	1E-14	leaf rust 10 disease-resistance locus receptor	*Punica granatum*	ATP binding; polysaccharide binding; protein serine/threonine kinase activity	-
SNP91	ABF73316.1	4E-14	clavata-like receptor	*Picea glauca*	ATP binding; protein kinase activity	-
SNP108	TKS08810.1	2E-41	DNAJ heat shock N-terminal domain-containing family protein	*Populus alba*	-	-
SNP133	ABG34278.1	6E-58	polygalacturonase	*Eucalyptus globulus*	Polygalacturonase activity	Carbohydrate metabolic process; cell wall organization
SNP151	XP_002320762.1	3E-13	peroxisomal membrane protein 11D	*Populus trichocarpa*	Identical protein binding	Peroxisome fission; peroxisome organization; regulation of eproxisome size
SNP159	ACJ09662.1	9E-45	putative calcium-dependent protein kinase, partial	*Cupressus sempervirens*	Kinase; transferase	-
SNP206	ABR15469.1	0.00	GDP-mannose pyrophosphorylase	*Pinus taeda*	Nucleotidyltransferase activity	Biosynthetic processes
SNP241	ATP71577.1	0.004	hypothetical protein	*Pinus pinaster*	-	-
SNP258	AJP06341.1	0.00	PIN2	*Pinus tabuliformis*	-	Auxin-activated signaling pathway; transmembrane transport
SNP273	AAQ63936.1	8E-15	cellulose synthase, partial	*Pinus radiata*	Cellulose synthase (UDP-forming); metal ion binding	-
SNP340	ABR15469.1	0.00	GDP-mannose pyrophosphorylase	*Pinus taeda*	Nucleotidyltransferase activity	Biosynthetic process
SNP350	AEX11975.1	2E-78	hypothetical protein	*Pinus taeda*	-	-

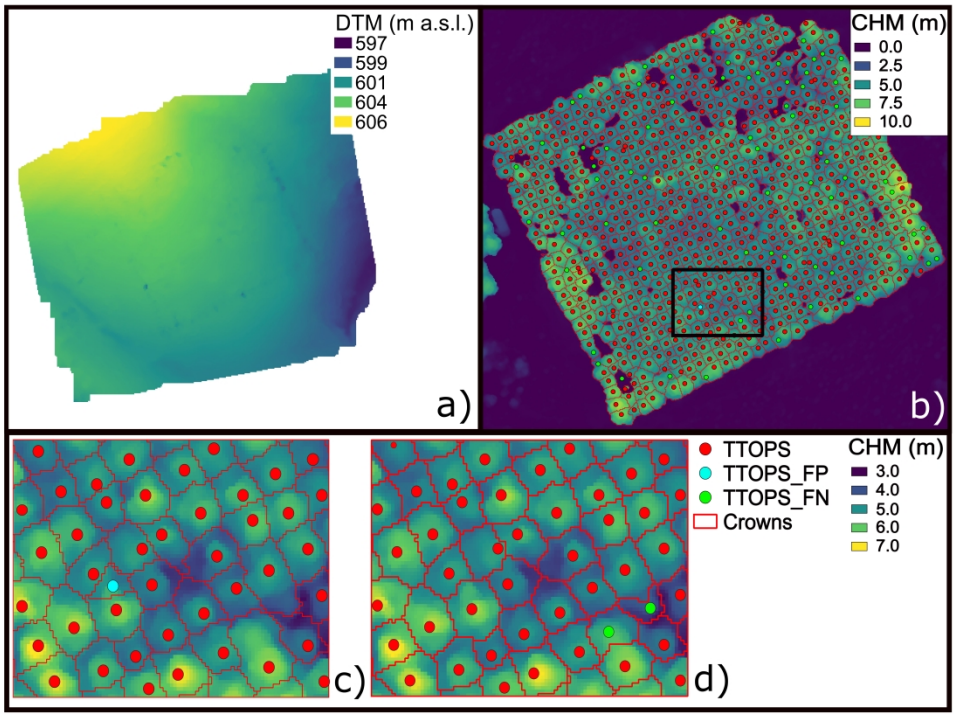
1019

Table 4 Results of the MLMM analysis. For each trait, the SNPs included as cofactor in the best fitting model are reported, as well as the percentage of variance explained (PVE) explained by the combination of SNPs.

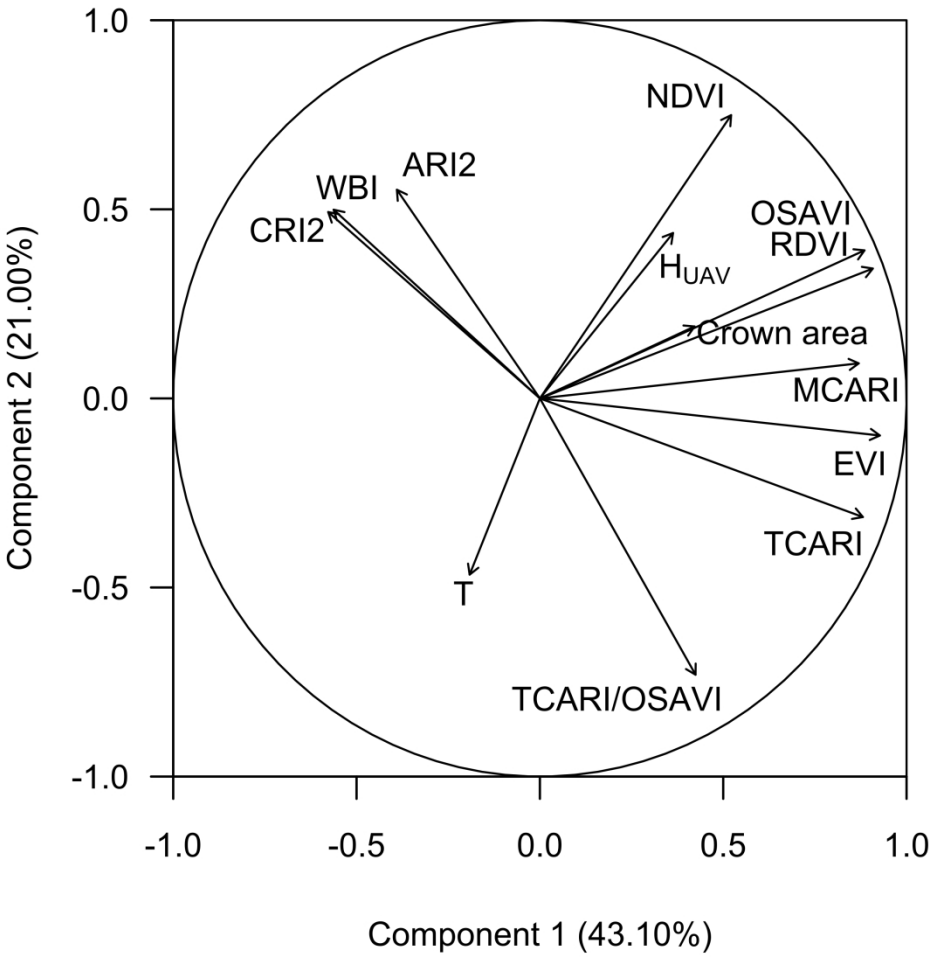
Trait	Indicator	SNPs as cofactors in the best fitting model	PVE (%)
Leaf area			
	NDVI	183	2.13
	OSAVI	183	3.05
	RDVI	108, 183	3.48
	EVI	16, 38, 106, 108, 222, 256, 264, 269, 326, 350	18.34
	MCARI	133, 222	5.17
	TCARI	38, 140, 222, 312, 350, 368	10.04
Photosynthetic pigments			
	TCARI/OSAVI	315	2.24
	ARI2	261	2.83
	CRI2	8	4.09
Water content			
	WBI	140, 156, 205, 216, 304, 312, 319, 343, 350	19.83
Canopy temperature			
	T	128, 140, 180, 215, 248, 258, 275, 319, 375	16.51
Growth traits			
	Crown area	21, 49	8.48
	H_{UAV}	18, 350	15.74



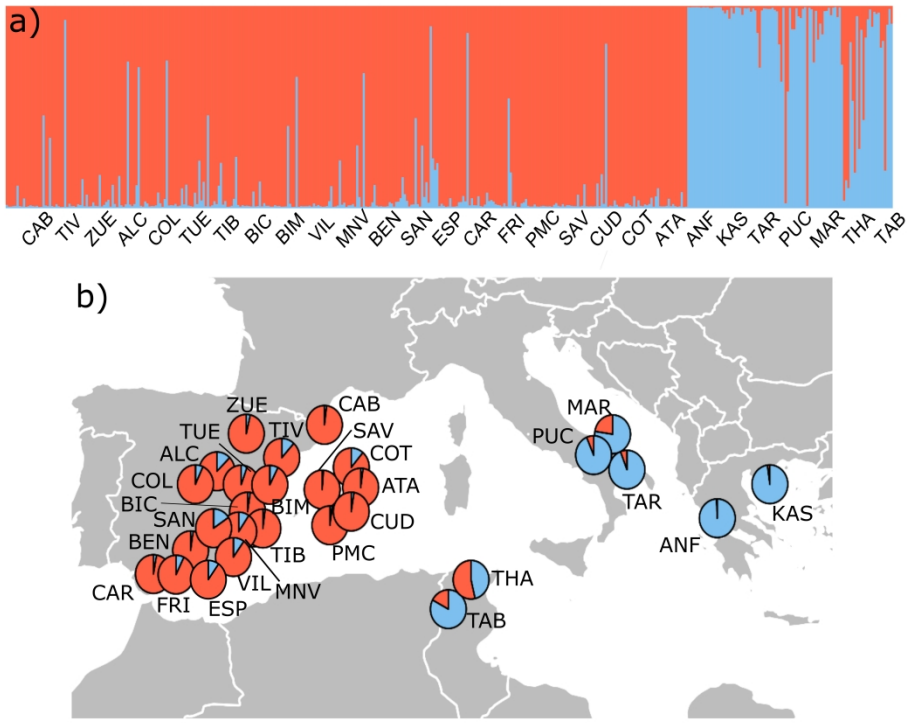
Flowchart showing the different steps followed to carry out the study of association genetics in P. halepensis using RGB, multispectral and thermal imagery as phenotyping tools.



a) Digital terrein model and b) canopy elevation model obtained through structure-from-motion analysis of RGB images. The correctly identified treetops (red dots), the false positive tree tops (i.e. no trees identified as trees; light blue dots) and the false negative treetops (i.e. trees not identified) are plotted on the canopy elevation model together with the segmented crown shapes (red polygons). An area of the trial (black square) has been enlarged to show the difference between crown shaped before (c) and after (d) the manual correction of the mis-idenfied treetops.



Plot of the loadings of the first two PCA axes describing the relations between UAV-derived growth traits, vegetation indices and canopy temperature. NDVI, Normalized Difference Vegetation Index; OSAVI, Optimised Soil Adjusted Vegetation Index; RDVI, Re-normalized Difference Vegetation Index; EVI, Enhanced Vegetation Index; MCARI, Modified Chlorophyll Absorption Reflectance Index; TCARI, Transformed Chlorophyll Absorption Ratio Index; CRI2 Carotenoid Reflectance Index 2; ARI2, Anthocyanin Reflectance Index 2; WBI Water Band Index; T, canopy temperature; HUAV, UAV-retrieved tree height.



a) Bar plot showing results of the assignment test with K=2 for the 375 genotyped individuals. Each individual is represented by a vertical line divided into two coloured segments representing the probabilities that the individual is assigned to the group comprising Western (in red) or Central-Eastern (in blue) Mediterranean populations. Population tags are reported in Table S1. b) Pie charts showing the percentage of assignment to the two groups in each population. The charts are plotted on the actual geographic coordinates of populations.